Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 253-264
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6254

Knowledge-Based Decision Support System for Determining Types of Agricultural Crops According to Soil Conditions

Ferry Wahyu Wibowo[1,*], Andi Sunyoto[1], Bayu Setiaji[1], Wihayati[2]

[1] Faculty of Computer Science, Informatics, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia
[2] Faculty of Information Technology, Computer Science, Satya Wacana Christian University, Salatiga, Indonesia
Email: [1,*]ferry.w@amikom.ac.id, [2]andi@amikom.ac.id, [3]bayusetiaji@amikom.ac.id, [4]wihayati@gmail.com
Correspondence Author Email: ferry.w@amikom.ac.id
Submitted: 13/11/2024; Accepted: 30/11/2024; Published: 30/11/2024

Abstract−Selecting the right crop for a particular land condition is one of the significant challenges in the agricultural sector. Each crop type has specific needs related to environmental factors such as soil type, pH, humidity, rainfall, and temperature. Mistakes in determining the appropriate crop type can result in decreased production, wastage of resources, and losses for farmers. This paper aims to determine the best model for use as a knowledge base to choose suitable plants for soil conditions. Machine learning algorithms were used to identify patterns of relationships between land conditions and the success of certain crop types to assist in selecting suitable crops and then made a knowledge-based decision support system. Algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), and k-Nearest Neighbor (k-NN) have been applied to solve this problem. In this paper, 30 experiments were conducted to test the stability of the model in determining suitable crops based on land conditions. The results of the experiments showed that the Support Vector Machine (SVM) has a more stable performance than other algorithms, with accuracy values of mean and standard deviation of 1 and 0, respectively.

Keywords: Agriculture; Crop; Decision Support System; Knowledge; Soil

1. INTRODUCTION

Agricultural land is one of the most essential natural resources for human life in food production, economy, and environmental sustainability. The existence of healthy and productive agricultural land also has a direct impact on food security, farmer welfare, and the socio-economic stability of the community [1]. Agricultural land produces various food products, such as rice, corn, wheat, vegetables, fruits, and other commodities. With the increasing world population, the need for food is increasing. Therefore, sufficient and productive agricultural land is essential for a sustainable food supply. Agricultural land provides food products and becomes a resource for industrial raw materials and renewable energy, such as corn, cotton, palm oil, sugar cane, and soybeans, which produce vegetable oil, sugar, textiles, and bioenergy. Agriculture is a source of livelihood for millions of farmers who depend on the output of agricultural land for their livelihoods [2]. Therefore, the availability and management of good agricultural land are crucial to support economic growth, reduce poverty rates, and increase income and community welfare, especially in rural areas. Fertile and reliable land conditions are also essential to maintain socio-economic stability in rural areas. When agricultural land is well managed, it can improve farmers' welfare, reduce poverty, and improve their access to education and health services [3]. Advanced agricultural technology and data-based management also provide great opportunities to increase the productivity of limited agricultural land.

Determining the type of crops based on soil conditions is a crucial step in agriculture that can affect the results and success of agricultural production [4]. Soil has physical, chemical, and biological characteristics that significantly affect plant growth [5]. Therefore, understanding good soil conditions for each crop type is essential to maximize agricultural productivity and sustainability [6]. Soil with high sand content has good drainage, but water retention and soil fertility tend to be low [7]. Plants suitable for sandy soil do not require high humidity, such as corn, soybeans, and secondary crops [8]. Soil with high clay content tends to be denser, stores water well and has high fertility, but drainage is less than optimal [9]. Plants suitable for clay soil are rice, soybeans, or vegetables such as cabbage and carrots. Meanwhile, soil containing a mixture of sand, silt, and clay has an ideal texture for many plants. Some plants suitable for clay soil include rice, corn, and chilies [10].

An Acidic Soil pH (pH < 6) is soil that is too acidic and can inhibit the availability of nutrients for plants [11]. Plants that are more resistant to acidic pH are plants such as rice, potatoes, tea, and coffee. Neutral Soil pH (pH 6 - 7) is a neutral pH with ideal soil conditions for most plants—suitable agricultural plants such as tomatoes, chilies, corn, and rice. Alkaline soil pH (pH > 7) in a high or alkaline pH is suitable for wheat, garlic, and several horticultural plants that are more resistant to alkaline soil conditions [12]. Soil with sufficient nutrient content is critical to support plant growth . Plants such as corn, green vegetables such as spinach and kale, and legumes need large amounts of nitrogen. Plants that produce fruit or tubers, such as tomatoes, potatoes, and carrots, need enough phosphorus and potassium in the soil to accelerate the flowering and ripening process of the fruit. Plants such as rice, sweet potatoes, and horticultural plants such as mustard greens and lettuce grow well in soil that has high humidity and can store water well. Some crops, such as corn, sorghum, and herbs like oregano and thyme, do better in well-drained soil and not too wet. Well-drained soil allows water to drain smoothly, avoiding puddles that can cause root rot [13]. Crops such as corn, tomatoes, and soybeans do well in

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 253-264
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6254

soil with good drainage. Meanwhile, soil that has poor drainage or tends to be waterlogged is suitable for crops resistant to water or puddles, such as rice and aquatic plants. In addition, the organic matter content in the soil contributes significantly to soil fertility. Soil with a high organic matter content, such as humus, is more fertile and can retain moisture and provide nutrients continuously. Plants such as vegetables and fruits will grow better in soil rich in organic matter.

A knowledge-based system (KBS) is a computer system that uses a knowledge base to solve problems or make decisions that usually require human expertise. KBS is designed to imitate experts' thinking and decision-making abilities in a particular field [14]. This system functions by storing and processing relevant information to help provide the right solution, similar to how humans solve problems using their knowledge. KBS works based on three main components: the knowledge base, inference engine, and user interface. The knowledge base contains a collection of facts, rules, or heuristics obtained from experts in a particular field. This knowledge can be in raw data, rules to follow, or more complex information for analysis and problem-solving. The inference engine is responsible for processing the information in the knowledge base. This engine makes decisions or suggests solutions based on existing facts and rules by using inference or reasoning techniques.

Meanwhile, the user interface allows users to interact with the system, provide input, and receive output. In many modern KBS, the user interface is also designed to be more intuitive, even for users who do not have technical knowledge. One type of KBS is decision support systems (DSS) [15]. DSS integrates various sources of knowledge to help decision-makers choose the best alternative [16]. This system focuses more on data processing and analytical models that enable business leaders to make better decisions based on data. Developing a knowledge-based system begins with collecting knowledge from various sources, especially experts in the relevant field. DSS can help farmers, land managers, and policymakers make more informed, data-driven decisions to improve agricultural productivity, efficiency, and sustainability [17]. Computer-based systems that allow users to make complex decisions using data, analytical models, and related knowledge. DSS integrates various information and resources in agriculture, such as weather data, soil conditions, cropping patterns, and market prices, to advise farmers or policymakers [18]. In general, the main objectives of DSS in agriculture are to improve operational efficiency, optimize agricultural output, minimize risk of losses, and support the long-term sustainability of food production. The expectation obtained when relying on DSS is that it can provide the right decisions in proper crop management based on data parameters caught by sensors. DSS helps farmers choose crops with optimal yield potential and reduce losses due to crop failure [19]. DSS also allows farmers to optimize using fertilizers, water, and pesticides. This system can calculate the proper dosage according to soil and crop conditions, reducing waste and negative environmental impacts. DSS can help farmers plan planting and harvesting times more precisely by exploiting weather data and crop yield prediction models [20]. Maximizing yields and minimizing losses due to bad weather or market fluctuations is very important. Agricultural DSS can help farmers manage risks associated with climate change, pests, plant diseases, or changes in market prices. By using recorded and predictive data, this approach provides a more pleasing understanding of potential risks and mitigation steps that it can take. DSS also supports sustainable agricultural practices by providing information on more environmentally friendly methods, such as using organic fertilizers or chemical-free farming techniques [21]. This system helps farmers understand and implement more environmentally friendly processes. Considering various environmental factors, this system helps farmers choose suitable crops to plant. One method that can be used in agricultural DSS is classification with labels such as rice, maize, mung bean, watermelon, and muskmelon. By utilizing classification, this DSS will provide recommendations for the types of crops most suitable for an area's land and climate conditions, which will help farmers make better decisions and increase their agricultural yields.

This paper is organized to make it easier for readers to understand the research conducted. After the Introduction section, which provides an overview of the topic and objectives of the research, the paper is divided into several main sections. In this paper, these sections consist of Research Methodology, Results and Discussion, and Conclusions. Each section has a vital role in conveying information systematically and in-depth. The research methodology explains in detail how the research was conducted. It includes the research design, the sample or object studied, data collection techniques, and the tools and methods of analysis used. A clear explanation of the methodology is essential so that the research can be replicated or well understood by other readers. The following section presents the results and discussion, which presents the main findings of the research that has been conducted. It is accompanied by tables, graphs, or diagrams to support the explanation, presenting facts and data found during the research. The last section of the scientific paper is the conclusion, which summarizes the main findings and implications of the research that has been conducted. In this section, the author draws conclusions based on the results and discussions that have been presented previously.

2. RESEARCH METHODOLOGY

Decision Support System (DSS) is a computer-based system that provides support in decision-making by processing and analyzing relevant data [22]. In agriculture, DSS aims to help farmers, farm managers, or governments make more informed decisions about things like selecting commodities to be planted, applying

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 253-264
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6254

cultivation techniques, and the right time to plant or harvest. Classification in DSS is the process of grouping data into specific categories based on certain characteristics or attributes. In this case, classification can be used to identify or predict the types of plants most suitable to be planted in a location based on factors such as soil content of nitrogen, phosphorus, potassium, pH, rainfall, temperature, and humidity. The process flow of this research is shown in Figure 1.

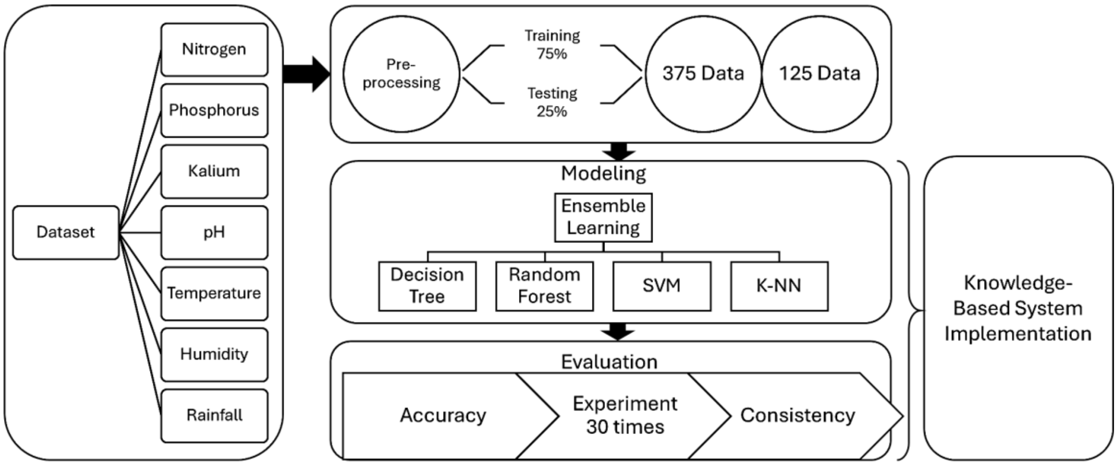


Figure 1. Flow Diagram of The Knowledge-Based Decision Support System for Determining Agriculture Crops

The dataset used was obtained from a 7-in-1 soil sensor, which measures various vital parameters related to soil conditions. This dataset was then adjusted to agronomic knowledge about the types of plants suitable for the soil conditions of the agriculture fields. Five hundred data were collected and divided into two parts: 375 data for the training process and 125 data for testing. The method implemented in this study is ensemble learning, which are Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). The learning process is not carried out only once but is carried out 30 times to test the consistency of the model.

2.1 Classification Parameters

The classification process in agricultural DSS to select the right plants involves several stages. The first step in this system is the collection of relevant data. The data used can include various information such as:

1. Soil conditions: soil pH, soil texture, nutrient content.
2. Climate: average temperature, rainfall, air humidity.
3. Topography: land slope, height above sea level.
4. Socio-economic conditions: market type, demand for agricultural products, and resource needs.

This paper only deals with information on soil conditions or nutrient content such as nitrogen, phosphorus, and potassium, as well as temperature, humidity, soil pH, and rainfall. The labels used in this classification refer to the types of plants that Indonesian farmers generally cultivate. Each type of plant has different needs and characteristics, so choosing plants suitable for local conditions is essential. Information about each label in this paper, namely:

1. Rice is Indonesia's leading food crop, requiring saturated soil conditions and sufficient rainfall.
2. Maize is a food crop that is more flexible than rice because it can grow in various types of soil and climates. Maize requires warm temperatures and moderate rainfall.
3. Mung Beans are legumes tolerant to dry soil and hot temperatures. This plant is more suitable for planting in areas with low rainfall and rather dry soil.
4. Watermelon is a seasonal fruit that requires a hot climate and fertile soil with good drainage.
5. Muskmelon requires conditions like those of watermelon, namely a warm climate and loose soil with good drainage.

The data used in this paper amounted to 100 for each classification label, for a total of 500.

2.2 Preprocessing

After the data is collected, it must be cleaned and prepared so that it is ready for processing. This data processing includes several steps, such as:

1. Dealing with missing or incomplete data.
2. Normalizing or standardizing the data to ensure each feature (such as soil temperature or pH) has a similar scale.
3. Feature selection to select the most relevant variables to the classification.

These data are separated into training and testing data with a ratio of 75% and 25%, respectively. Thus, the data applied were 375 and 125 data.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 253-264
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6254

2.3 Modeling

Once the data is ready, the next step is to choose the correct classification algorithm. The algorithms used in agricultural classification in this paper are:

a. Decision Tree, which functions to make decisions based on a series of questions that divide the data. It can be written as pseudocode shown on Algorithm 1.

Table 1. Pseudocode Decision Tree Algorithm

Algorithm 1. Decision Tree

```
IF data ← Ɵ
   RETURN MajorityClass(data)
IF attributes ← Ɵ
   RETURN MajorityClass(data)
BestAttribute ← ChooseBestAttribute(data, attributes)
TreeNode ← CreateNode(bestAttribute)
AttributeValues ← GetPossibleValues(BestAttribute)
FOR value ← attributeValues
   Subset = FilterDataByAttributeValue(data, BestAttribute, value)
   IF subset ← NOT Ɵ
       Subtree ← DecisionTreeLearning(subset, RemoveAttribute(attributes, BestAttribute))
       AttachSubtreeToNode(TreeNode, value, Subtree)
   ELSE
       Leaf ← MajorityClass(data)
       AttachLeafToNode(treeNode, value, Leaf)
RETURN TreeNode
```

b. Random Forest, which combines many decision trees to increase accuracy. Its working principle is shown in algorithm 2.

Table 2. Pseudocode Random Forest Algorithm

Algorithm 2. Random Forest

```
forest = []
FOR i = 1 to nTrees
    BootstrapData = BootstrapSample(trainingData)
    Tree = BuildDecisionTree(bootstrapData, mFeatures)
    AddTreeToForest(forest, tree)
RETURN forest
Function BuildDecisionTree(data, mFeatures)
    IF StoppingConditionMet(data)
          RETURN CreateLeafNode(data)
featuresToConsider = RandomlySelectFeatures(data, mFeatures)
bestFeature = ChooseBestFeature(data, featuresToConsider)
treeNode = CreateNode(bestFeature)
subsets = SplitData(data, bestFeature)
FOR subset ← subsets
    childNode = BuildDecisionTree(subset, mFeatures)
    AttachChildNode(treeNode, childNode)
RETURN treeNode
predictions = []
FOR tree ← forest
     prediction = PredictWithTree(tree, testData)
     AddPredictionToResults(predictions, prediction)
finalPrediction = MajorityVoting(predictions)
RETURN finalPrediction
```

c. Support Vector Machine (SVM) is an algorithm that functions to find the best line or hyperplane by distinguishing categories in the data. Its working principle is shown in algorithm 3.

Table 3. Pseudocode Support Vector Machine Algorithm

Algorithm 3. Support Vector Machine

```
Kernel ← linear, polynomial, RBF
Regularization C
```

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 253-264
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6254

Algorithm 3. Support Vector Machine

Matrices ← (Kernel function ← $K(x_i, x_j)$)
Minimize $0.5 * ||w||^2 + C * \Sigma(\xi_i)$
Subject to: $y_i * (w \bullet x_i + b) \geq 1$
ξ_i for each i, ξ_i: slack variable & b: bias
$f(x) = w \bullet x + b$
Find f(x')
IF $f(x') \geq 0$
    x' ← positive
IF $f(x') < 0$
    x' ← negative
RETURN finalPrediction

d. k-Nearest Neighbors (k-NN) has a principle of data classification based on its proximity to other data that has been grouped. Its working principle is shown in algorithm 4.

Table 4. Pseudocode k-Nearest Neighbors Algorithm

Algorithm 4. k-Nearest Neighbors

```
Nearest Neighbors ← k
data ← x'
centroid ← X
FOR i in data
    Find a distance between x' & each centroid
    Save a distance with class label of each data
Sort all data
Select sampel k with closest to x'
Find majority of Nearest Neighbors k
RETURN finalPrediction
```

This classification model is trained using processed data to identify patterns or relationships between environmental variables and suitable plant types.

2.4 Model Evaluation

Once the model is trained, it must be tested using test data. This ensures that the model not only memorizes the training data but can also accurately classify new data. The model evaluation uses accuracy, but to see the consistency of the model, 30 experiments are needed, which are then analyzed statistically. It is to get the best model.

2.5 System Implementation

Once the model is ready and tested, the system can provide recommendations to farmers. Based on the input data on land and climate conditions, the system will suggest the types of crops that are most suitable to be planted. It will help farmers choose suitable crops to increase their yields.

3. RESULT AND DISCUSSION

The dataset used in this study is shown in Figure 2. This dataset was obtained through a 7-in-1 soil sensor that measures various important parameters, such as humidity, temperature, pH, and soil nutrient levels. The labels indicate the suitability of the soil for multiple types of plants.

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

Figure 2. Some examples of datasets used to determine the parameters used to determine suitable crops

This paper encodes the kinds of plants such as maize, mungbean, muskmelon, rice, and watermelon into 0, 1, 2, 3, and 4, respectively. It is shown in Table 1.

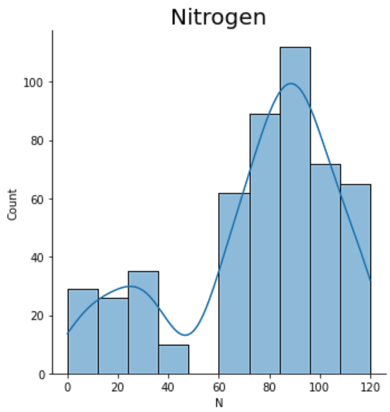
Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 253-264
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6254

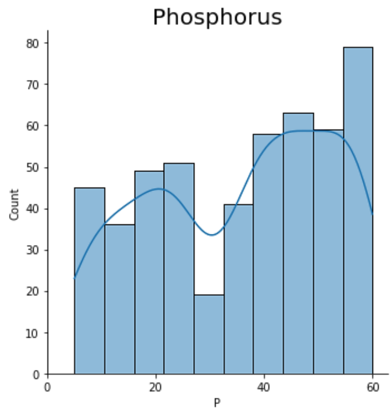
Table 1. Label encoding of crop names

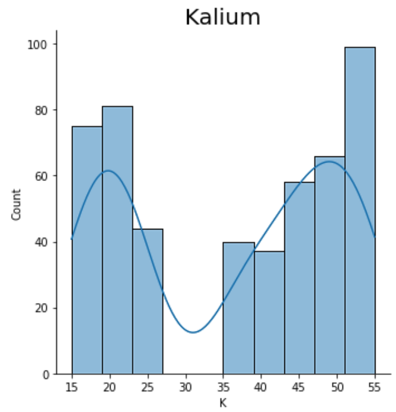
No	Crop Name	Label Encoding
1	Maize	0
2	Mungbean	1
3	Muskmelon	2
4	Rice	3
5	Watermelon	4

Encoding data into numeric format efficiently prepares categorical data before being processed by a machine learning model. Assigning unique values to each category can help the model identify differences between plant types in a way that is easy to understand and compute.

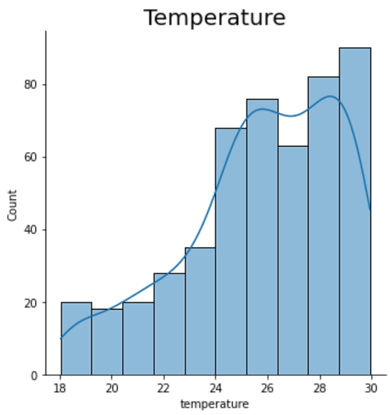
A histogram can show how spread out or variable the data is. A wide spread of data can be seen from a broader range histogram, while data consistently in a small interval tends to show a narrower spread. The histogram of each parameter used in training the machine learning model is shown in Figure 3.

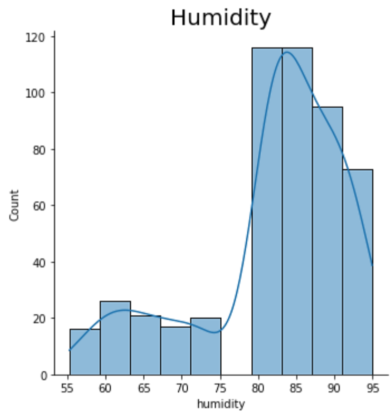


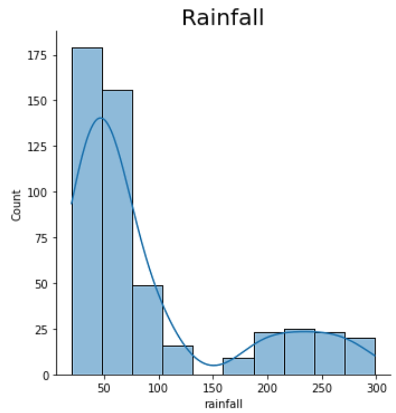




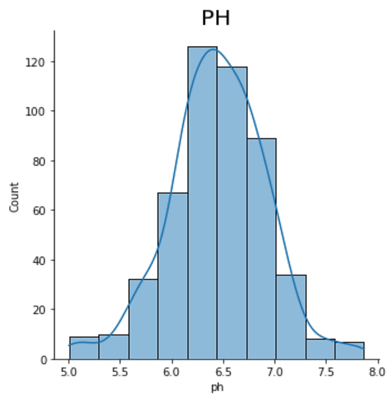
(a) Histogram of Nitrogen Dataset (b) Histogram of Phosphorus Dataset (c) Histogram of Kalium Dataset







(d) Histogram of Temperature Dataset (e) Histogram of Humidity Dataset (f) Histogram of Rainfall Dataset

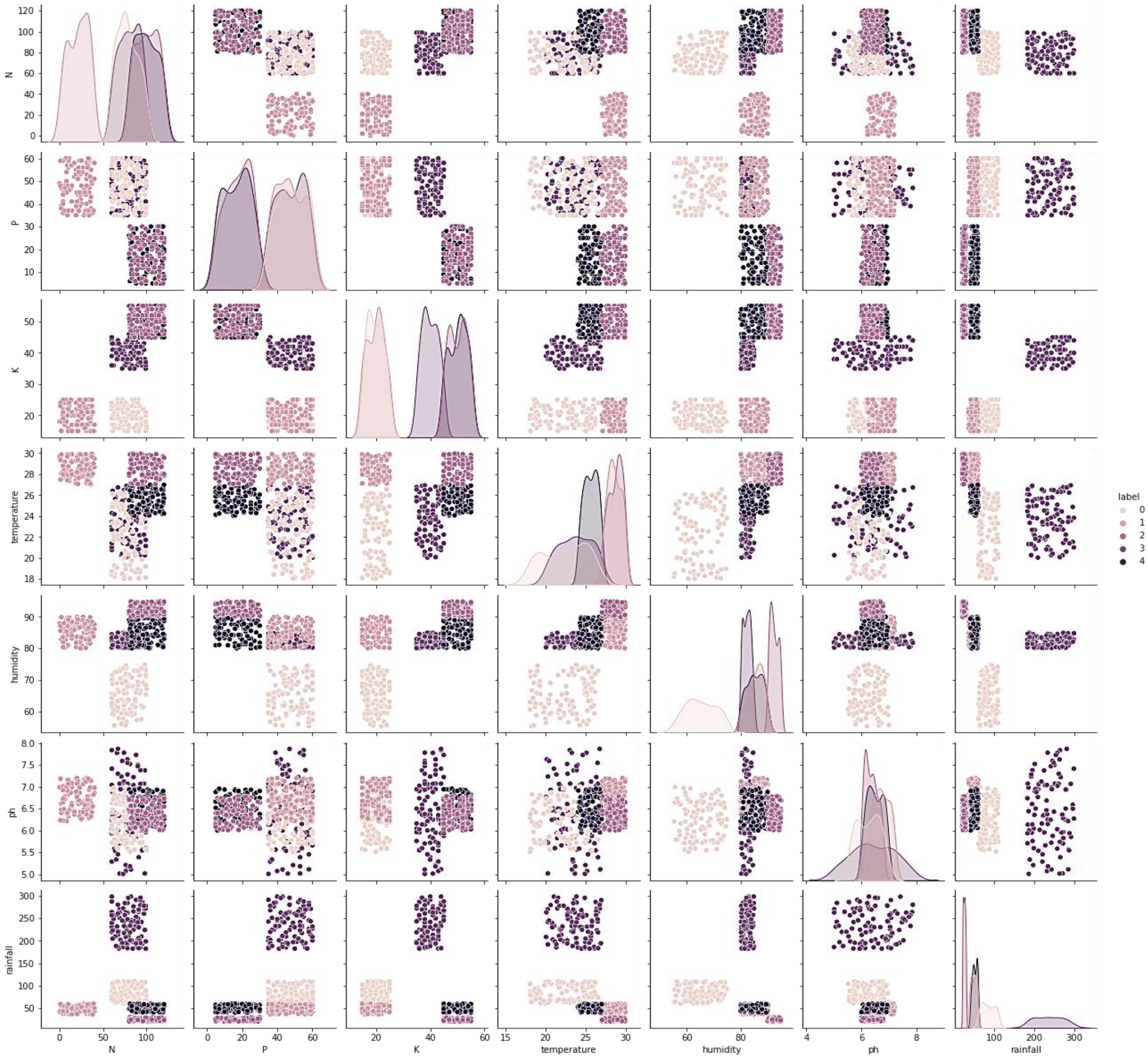


(f) Histogram of pH Dataset

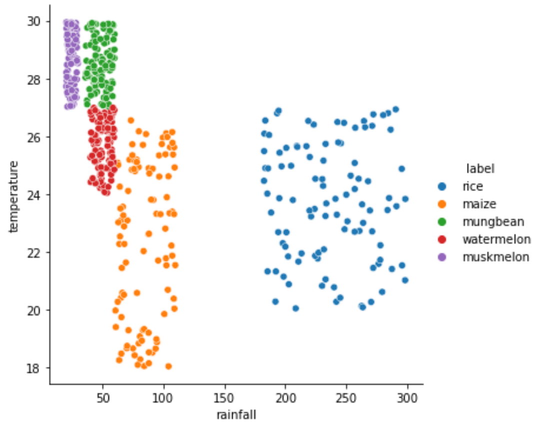
Figure 3. Histograms of Each Parameter Uses Ten Bins for Each Parameter

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 253-264
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6254

In Figure 3, the pH dataset's histogram graph is almost symmetrical compared to the others, all of which have bimodal graphs. A histogram is also helpful for comparing two data groups and can see patterns of change or differences in distribution between the two periods. If the histogram is bell-shaped, it indicates that the data tends to be normally distributed. If there are bins with very low frequencies on the histogram's left or right ends, this shows very low or very high data. By visualizing the frequency of occurrence of data in specific bins, histograms provide insight into the essential characteristics of the data that can be used for decision-making and further analysis. A comparison between the two parameter data is shown in Figure 4.



(a) Comparison Between Parameters

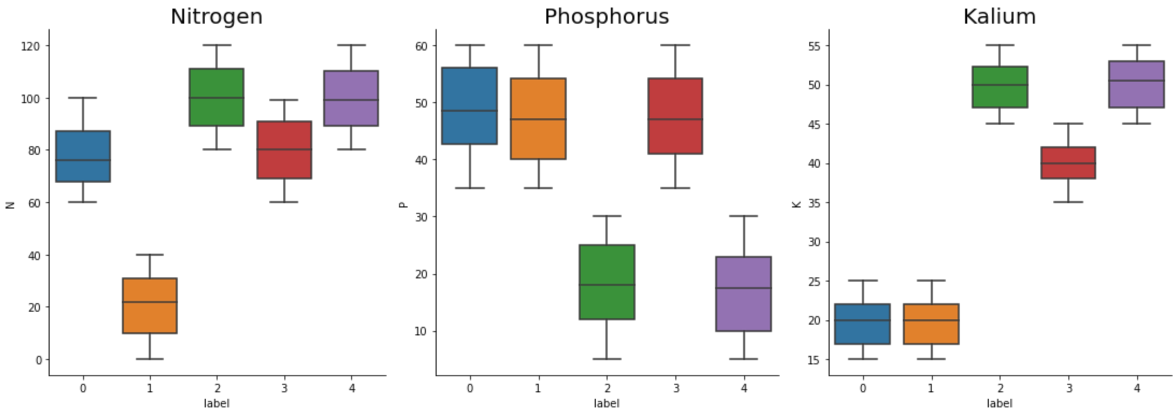


(b) Comparison Between Rainfall vs Temperature on Each Label

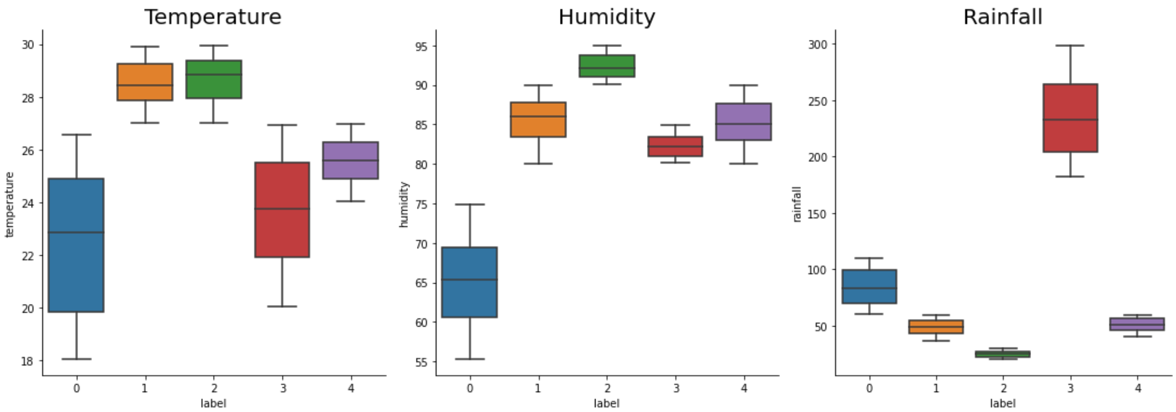
Figure 4. Comparison Between Parameters on Each Label

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 253-264
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6254

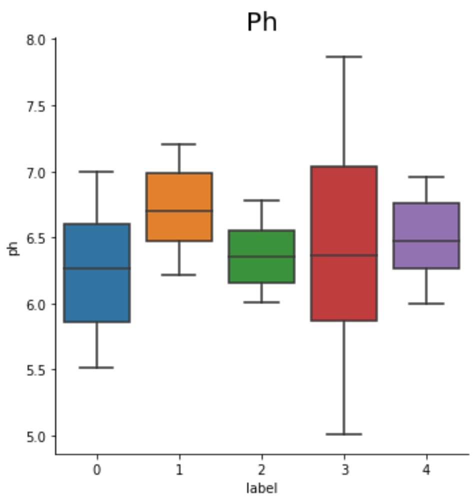
In addition, a boxplot representation for summarizing data distribution and providing an overview of the spread, the position of the data, and potential outliers. Boxplots present data in a simple yet informative form, focusing on five main data measures, namely the minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum. Boxplots of each dataset parameter are shown in Figure 5.



(a) Boxplot of Nitrogen Dataset (b) Boxplot of Phosphorus Dataset (C) Boxplot of Kalium Dataset



(d) Boxplot of Temperature Dataset (e) Boxplot of Humidity Dataset (f) Boxplot of Rainfall Dataset



(d) Boxplot of Temperature Dataset

Figure 5. Boxplots of Each Dataset Parameter on Each Label

From the boxplot, it can be seen that the distribution of labeling has varying means and distributions. For each algorithm, 30 tests were conducted. The mean and standard deviation (Std) of each test result were then calculated to evaluate the average performance and consistency of the model. The mean indicates the average accuracy score, or other metrics generated from each experiment, while the standard deviation shows the scores' variable or consistency. Table 2 shows the results of this experiment.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 253-264
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6254

Table 2. The results of 30 experiments to see the consistency of a model through accuracy

No	Model			
	Decision Tree	Support Vector Machine	Random Forest	k-Nearest Neighbors
1	1	1	0.992	1
2	0.992	1	0.992	0,992
3	1	1	1	1
4	0.992	1	1	1
5	1	1	1	1
6	1	1	1	0,992
7	0.992	1	0.992	0,992
8	1	1	1	1
9	0.992	1	1	1
10	1	1	1	0,992
11	0.992	1	0.992	0,992
12	1	1	1	1
13	0.992	1	1	1
14	1	1	0.992	1
15	0.992	1	0.992	0,992
16	1	1	1	1
17	0.992	1	1	1
18	1	1	1	1
19	1	1	1	0,992
20	0.992	1	0.992	0,992
21	1	1	1	1
22	0.992	1	1	1
23	1	1	0.992	1
24	1	1	1	0,992
25	0.992	1	0.992	0,992
26	1	1	1	1
27	0.992	1	1	1
28	1	1	1	0,992
29	0.992	1	1	1
30	1	1	0.992	1
Mean	0.996533	1	0.997333	0.997067
Std	0.004032	0	0.003836	0.003921

The mean of Support Vector Machine (SVM) has the highest average score compared to the other algorithms, indicating that this method is more accurate in classifying data over 30 experiments. The standard deviation of SVM is low, indicating good consistency. In other words, the performance of SVM is relatively stable and does not change much between experiments. The mean of Random Forest is slightly lower than SVM but still higher than k-Nearest Neighbors (k-NN) and Decision Tree. The Std of Random Forest has a lower standard deviation than k-NN and Decision Tree, which means that this model is relatively consistent in its performance, although not as high as SVM.Best Performance on SVM: The experimental results show that SVM has the highest mean and low standard deviation, indicating that SVM is accurate and consistent in its performance. This algorithm separates data around the hyperplane, primarily when the data is distributed in a pattern that a large margin can separate. This stability makes SVM a good choice for cases where accuracy and consistency are crucial. Random Forest is Superior to k-NN and Decision Tree: Random Forest has a higher mean than k-NN and Decision Tree, indicating that this ensemble technique can utilize many decision trees to improve accuracy. Combining many trees and a Random Forest can reduce overfitting, which is often a problem in a single Decision Tree. In addition, the low standard deviation of Random Forest indicates good stability, making this model reliable in repeated experiments.

This experiment used hyperparameter tuning on these algorithms. The technique used for tuning is Grid Search, which performs an exhaustive search on all possible parameter combinations to find the best configuration that provides optimal performance for each model. After the tuning process, the results showed an increase in performance for all models compared to the previous experiment without tuning. The results of hyperparameter tuning for each model are presented in Table 3, which includes the optimal parameters successfully found through Grid Search.

Table 3. Optimal parameters of the models

No	Model	Parameter
1	Decision Tree	splitter: best

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 253-264
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6254

No	Model	Parameter
2	SVM	c: 1, kernel: radial basis function (rbf)
3	Random Forest	n_estimators: 5
4	k-Nearest Neighbors	n_neighbors: 5, weights: uniform

The tuning results show that hyperparameter tuning with Grid Search can significantly improve the model's performance. This improvement is especially visible in the SVM and Random Forest models, which both have higher average accuracy and lower standard deviation after tuning. In this experiment involving classifying fruit types, specifically muskmelon and watermelon, the accuracy results were 0.992 or 99.2%. Although this level of accuracy is very high, several cases of inaccuracy or prediction errors are interesting to analyze further. One of the main factors contributing to this error is the similarity of variables between muskmelon and watermelon, which causes the model to need help distinguishing between the two classes, as shown in Figure 6.

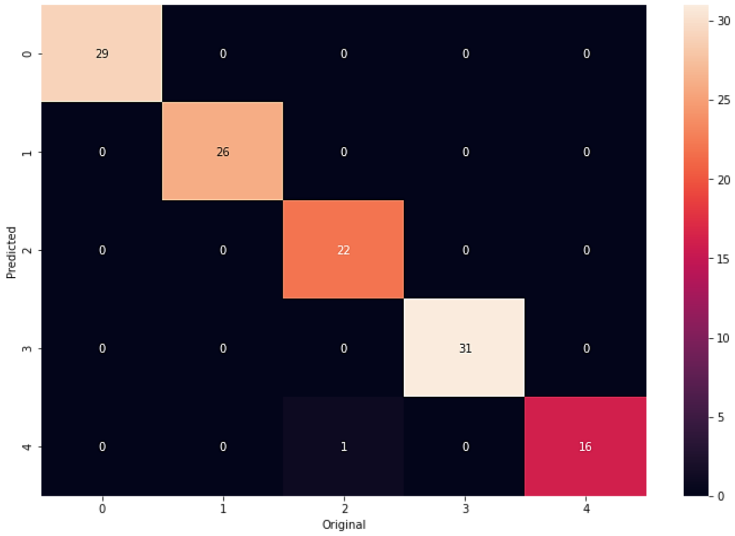


Figure 6. Heatmap of Classification Model Evaluation

In Figure 6, the distribution of muskmelon and watermelon variables shares several characteristics. This visualization shows the areas where the variables of the two types of fruit overlap, which is a source of prediction errors for the model. So, the SVM model is the implementation result used in determining agricultural crops suitable for several condition parameters of the selected land.

4. CONCLUSION

A classification-based decision support system (DSS) with labels for maize, mungbeans, muskmelon, rice, and watermelon can be a handy tool for farmers in selecting the most suitable crops for their environmental conditions. Using relevant data and appropriate classification algorithms, this DSS can help increase agricultural productivity, reduce the risk of crop failure, and support more efficient management of natural resources. The Ensemble Learning method for determining crops can produce fast and precise decision computations. The reliability of the technique is because the parameters used have significant differences so that the machine learning algorithm can obtain the pattern. Decision Tree breaks down data based on certain features, forming a decision tree structure. Each branch on the tree represents a decision that leads to a class or category prediction. In the context of plant selection, Decision Tree can be used to determine plant types based on land factors hierarchically. Random Forest is generally more powerful than Decision Tree because it reduces overfitting and provides more stable results. Support Vector Machine (SVM) is an algorithm that separates data into different classes by finding a hyperplane that maximizes the margin between classes. In plant selection, SVM can better help separate suitable and unsuitable plant types, especially if the data is not linearly distributed. K-Nearest Neighbor (k-NN) classifies new samples based on their similarity to existing samples. K-NN can predict the most likely suitable plants by looking for plants on similar land. In this paper, 30 experiments were conducted to test the stability of the model in determining suitable crops based on land conditions. The experimental results show that Support Vector Machine (SVM) has a more stable performance than other algorithms. The best results were obtained by SVM modeling with an average accuracy value and standard deviation of 1 ± 0. Stability is measured based on the model's ability to produce consistent prediction results over several trials. Factors that influence this stability come from the nature of SVM, which is very effective in handling data with large margins

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 253-264
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6254

and separating classes and is suitable for land conditions with unique characteristics. In the future, applying this technology can be essential in realizing a more sustainable and profitable agriculture for farmers.

ACKNOWLEDGMENT

The research of this paper is supported by research funds from the Ministry of Education, Culture, Research and Technology through research contract number 0459/E5/PG.02.00/2024 and derivative contract number 107/E5/PG.02.00.PL/2024 entitled Optimization of Communication Between Wireless Sensor Network Agents for Coastal Agricultural Land Using the Internet of Things.

REFERENCES

[1] I. M. Mehedi, M. S. Hanif, M. Bilal, M. T. Vellingiri and T. Palaniswamy, "Remote Sensing and Decision Support System Applications in Precision Agriculture: Challenges and Possibilities," in IEEE Access, vol. 12, pp. 44786-44798, 2024, doi: 10.1109/ACCESS.2024.3380830.

[2] P. Wolo, S. Mulyana, And R. Wardoyo, “Decision Support System For Suitability Of Horticultural Agricultural Plant Types With Land Conditions Using Interpolation Profile Matching,” in JUITA: Jurnal Informatika, vol. 11, no. 1, pp. 133-143, 2023.

[3] H. Yang et al., "Integrating Multidimensional Feature Indices and Phenological Windows for Mapping Cropping Patterns in Complex Agricultural Landscape Regions," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 7448-7459, 2024, doi: 10.1109/JSTARS.2024.3379216.

[4] J. Babber, P. Malik, V. Mittal and K. C. Purohit, "Analyzing Supervised Learning Algorithms for Crop Prediction and Soil Quality," 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2022, pp. 969-973, doi: 10.1109/ICCMC53470.2022.9753798.

[5] Y. Andrayani, M. B, And A. Franz, “Decision Support System For Selection Of Productive Land In Corn Using The Smart Method,” in Tepian, Vol. 4, No. 1, Mar. 2023, doi: 10.51967/Tepian.V4i1.1644.

[6] K. P. Shah, S. Narayan Sah, K. P. Dharmaraj and K. Dinesh Kumar, "A Comprehensive Analysis of Machine Learning Algorithms for Suitable Crop Prediction in Agriculture," 2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS), Coimbatore, India, 2024, pp. 63-68, doi: 10.1109/ICC-ROBINS60238.2024.10533902.

[7] A. A. AlZubi and K. Galyna, "Artificial Intelligence and Internet of Things for Sustainable Farming and Smart Agriculture," in IEEE Access, vol. 11, pp. 78686-78692, 2023, doi: 10.1109/ACCESS.2023.3298215.

[8] L. S. Kyrgiakos, G. Kleftodimos, M. D. Kremantzis, G. Vlontzos, And P. M. Pardalos, “Assessing Efficiency Differences In A Common Agriculture Decision Support System - A Comparative Analysis Between Greek And Italian Durum Wheat Farms -,” in International Journal On Food System Dynamics, vol. 14, no. 1, pp. 22–45, 2023, doi: 10.18461/Ijfsd.V14i1.E2.

[9] K. Chicaiza, R. X. Paredes, I. M. Sarzosa, S. G. Yoo and N. Zang, "Smart Farming Technologies: A Methodological Overview and Analysis," in IEEE Access, vol. 12, pp. 164922-164941, 2024, doi: 10.1109/ACCESS.2024.3487497.

[10] M. Kandan, G. S. Niharika, M. J. Lakshmi, K. Manikanta and K. Bhavith, "Implementation of Crop Yield Forecasting System based on Climatic and Agricultural Parameters," 2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT), Visakhapatnam, India, 2021, pp. 207-211, doi: 10.1109/ICISSGT52025.2021.00051.

[11] M. Sophocleous et al., "A Stand-Alone, In Situ, Soil Quality Sensing System for Precision Agriculture," in IEEE Transactions on AgriFood Electronics, vol. 2, no. 1, pp. 43-50, March-April 2024, doi: 10.1109/TAFE.2024.3351953.

[12] X. Lu, Q. Chen, Y. Ao, L. Li, And C. Tu, “Differentiation Of Mountainous Soil Ph And Its Influencing Factors In Guizhou Province, China,” in Pol J Environ Stud, vol. 33, no. 2, pp. 1277–1285, 2024, doi: 10.15244/Pjoes/172765.

[13] A. Bhermana And Susilawati, “Decision Support System Based On Database System Of Genetic Resources For Central Kalimantan Local Crops To Develop Ex Situ And In Situ Conservation,” In Iop Conference Series: Earth And Environmental Science, Institute Of Physics Publishing, vol. 482, pp. 012012, May 2020. doi: 10.1088/1755-1315/482/1/012012.

[14] D. Sharma And S. Kumar, “Knowledge Based Decision Support System In Steel Industries,” in Iop Conf Ser Mater Sci Eng, vol. 1116, no. 1, pp. 012083, 2021, doi: 10.1088/1757-899x/1116/1/012083.

[15] Ramadiani, B. Ramadhani, Z. Arifin, M. L. Jundillah, And A. Azainil, “Decision Support System For Determining Chili Land Using Weighted Product Method,” in Bulletin Of Electrical Engineering And Informatics, vol. 9, no. 3, pp. 1229–1237, Jun. 2020, doi: 10.11591/Eei.V9i3.2004.

[16] S. A. Bhat and N. -F. Huang, "Big Data and AI Revolution in Precision Agriculture: Survey and Challenges," in IEEE Access, vol. 9, pp. 110209-110222, 2021, doi: 10.1109/ACCESS.2021.3102227.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 6, No. 1, November 2024, Page 253-264
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v6i1.6254

[17] N. Rasheed, S. A. Khan, A. Hassan and S. Safdar, "A Decision Support Framework for National Crop Production Planning," in IEEE Access, vol. 9, pp. 133402-133415, 2021, doi: 10.1109/ACCESS.2021.3115801.

[18] T. Arunprasath, Y. Akshaya, B. Sonasri, P. R. M, S. Sankaran and K. Ramaraj, "Improving the Soil Fertility Based on Soil Texture and Climate Change," 2023 International Conference on Next Generation Electronics (NEleX), Vellore, India, 2023, pp. 1-6, doi: 10.1109/NEleX59773.2023.10421232.

[19] M. Iqbal Dzulhaq, M. Khadafi, S. Ramdhan, K. Sabardiman, S. Rahayu, And M. Bucci Ryando, "Food Crop Rotation Decision Support System Using Fuzzy," in Technology, And Engineering (Jste), vol. 1, no. 2, 2021.

[20] C. Ji, Y. Andrashko, S. Biloshchytska and S. Tsiutsiura, "Conceptual Research Model of Developing the Decision Support System for Agriculture Under Uncertainty," 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), Nur-Sultan, Kazakhstan, 2021, pp. 1-5, doi: 10.1109/SIST50301.2021.9465888.

[21] H. Devaraj et al., "RuralAI in Tomato Farming: Integrated Sensor System, Distributed Computing, and Hierarchical Federated Learning for Crop Health Monitoring," in IEEE Sensors Letters, vol. 8, no. 5, pp. 1-4, May 2024, Art no. 5501604, doi: 10.1109/LSENS.2024.3384935.

[22] F. Khemakhem, H. Ellouzi, H. Ltifi and M. B. Ayed, "Agent-Based Intelligent Decision Support Systems: A Systematic Review," in IEEE Transactions on Cognitive and Developmental Systems, vol. 14, no. 1, pp. 20-34, March 2022, doi: 10.1109/TCDS.2020.3030571.